



Arbil

Local Metadata Manager

Peter Withers
The TLA Language Archive,
Max Planck Institute for Psycholinguistics Wundtlaan
1, 6525 XD, Nijmegen, The Netherlands
peter.withers@mpi.nl





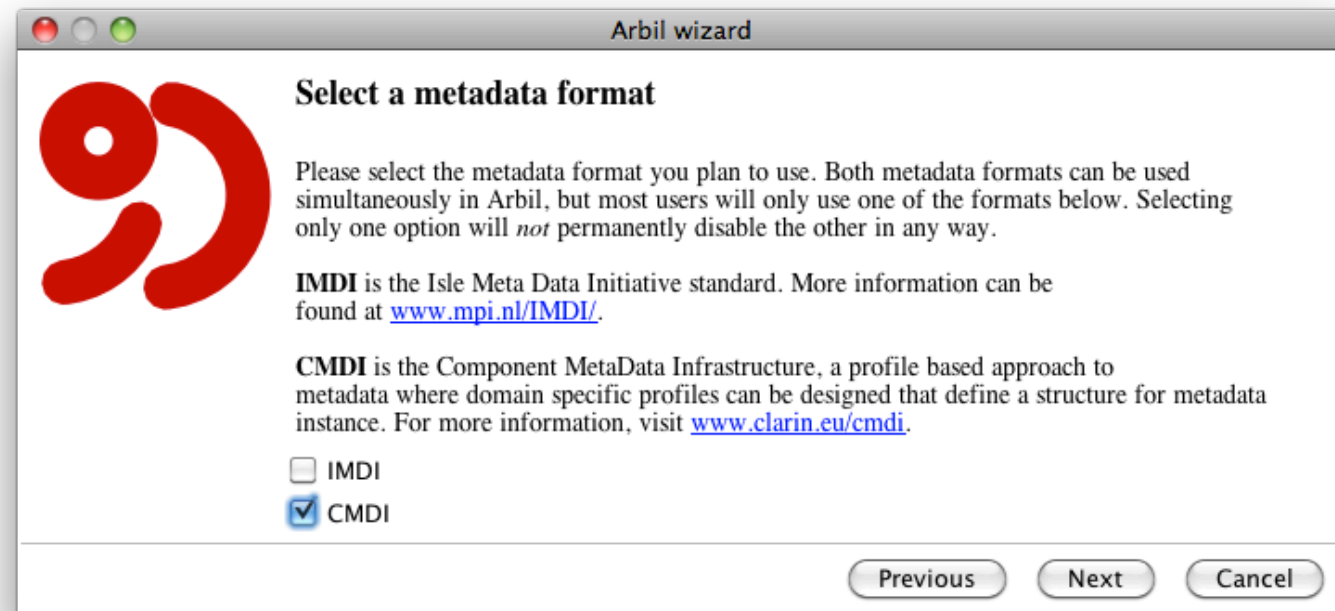
What Arbil Does

- ARBIL is an application for manually creating and organising metadata.
 - It is not designed to automate the metadata creation process.
 - There are other tools designed for more than 1000 or so metadata records.
- All editing within Arbil is done offline.
- Metadata structures can be created, viewed and edited.
- Multiple metadata files can be edited at the same time.
- The metadata to be arranged into corpus trees.
- The local metadata tree can be searched with multiple parameters.
- The data files can be launched in the associated applications such as ELAN or Media Players.
- Many metadata files can be bulk edited in a single table.
- With the exception of accessing the remote archive, all features are available offline such as in remote field sites.
- You can enter the data as it becomes available, there is no mandatory order.



Metadata Formats

- Both IMDI and Clarin metadata formats are supported.
- Other XML formats can potentially be supported by the use of custom templates or schema files.



Metadata Formats

IMDI (ISLE Meta Data Initiative) is a metadata format for linguistic data

- Has been around for about a decade
- Targeted towards multimodal / multimedia
- Has a single schema file with fixed metadata fields and optional key fields
- Is used for instance throughout the Language Archive in Nijmegen and associated archives



Clarín metadata is a flexible format with ISOCAT

- It is a more recent metadata format
- Each profile has its own schema files
- Developed as a part of the Clarín EU project
- Can be adapted to suit specific project needs
- The metadata fields and layout are customisable





The User Interface

Remote Corpus
View and import metadata from remote servers

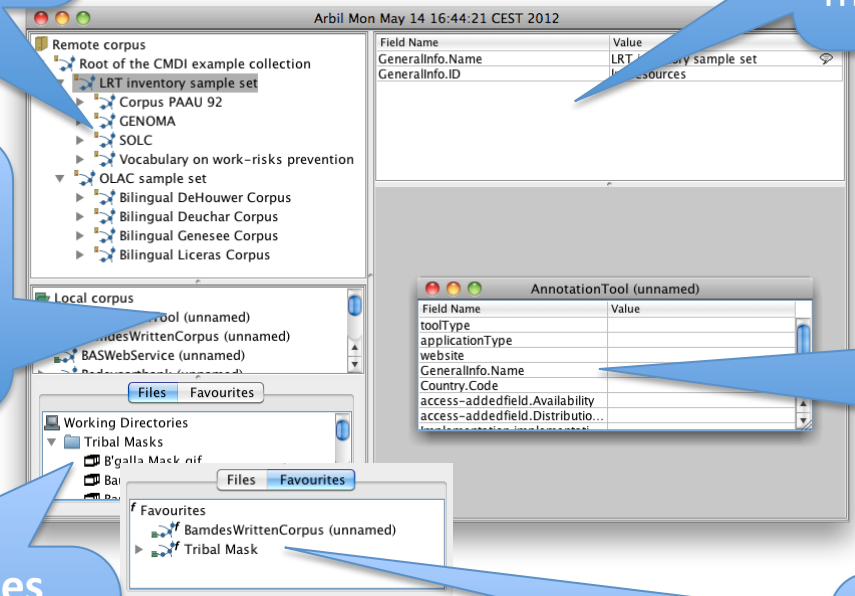
Local Corpus
All newly created metadata will be created here

Working Directories
Your data files can be browsed and associated with new metadata from here

Preview Table (optional)
The currently selected metadata.

Main Work Area
Multiple tables of metadata can be viewed and edited

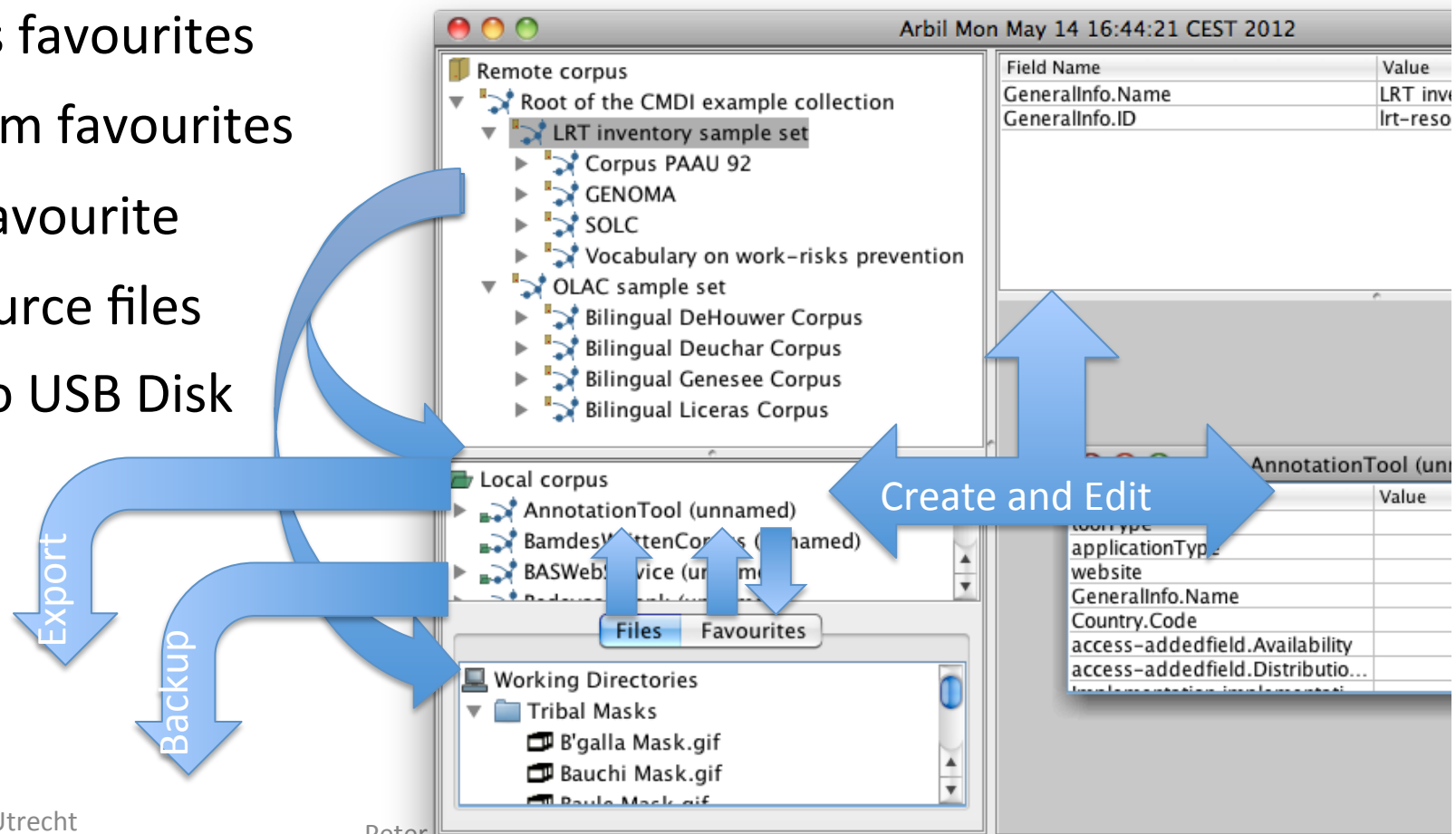
Favourites
Frequently used metadata is saved here for easy replication





Workflow

- Create and edit metadata
- Import for offline use
- Import as favourites
- Insert from favourites
- Save as favourite
- Add resource files
- Backup to USB Disk
- Export



Configuring ARBIL

- When Arbil is first started, you will be asked which metadata format to use.
- Depending on how you intend to use ARBIL you will want to select IMDI templates or Clarin profiles.
- You can change later in the templates & profiles dialogue.
- This can also be preconfigured by your system administrator.

The selected profile will be available via the add menu

Field Name	Value
toolType	
applicationType	
website	
GeneralInfo.Name	
Country.Code	
access-addedfield.Availability	
access-addedfield.Distribution	
Implementation.implementation	



Metadata Display

- The metadata is displayed in tables and trees, which allow an overview of the metadata and the ability to populate and update many metadata sections in bulk.

Remote corpus

- Root of the CMDI example collection
 - LRT inventory sample set
 - Corpus PAAU 92
 - GENOMA
 - ISO639 (2)
 - Catalan; Valencian
 - Spanish; Castilian
 - LrtCollectionDetails (unnamed)
 - Spain

Callouts:

- Metadata files are shown in the tree
- Multiple metadata files or subsections can be viewed in a single table
- metadata will be highlighted in red
- metadata are shown as separate subnodes

				LrtCo...	LrtCommon.Des
Corpus PAAU 92	Corpus PAAU 92	Institut Universitari de Li...	Written Corpus	(CV)	The electronic v
GENOMA	GENOMA	Institut Universitari de Li...	Written Corpus	(CV)	Bilingual written
SOLC	SOLC	Institut Universitari de Li...	Application / Tool	(CV)	An orthologic se
Vocabulary on work-risks...	Vocabulary on work-risks prevent...	Institut Universitari de Li...	Terminological Resou...	(CV)	An electronic ve



Table Views

- When multiple metadata files or subsections are viewed in a single table, they are each shown in a separate row.
- When a table shows a single metadata file or subsection, it is shown in the long view.
- Individual fields can be edited in the long field view, which allows each field to be edited sequentially.

The screenshot shows two windows from a software application. The left window, titled 'Corpus PAAU 92', displays a table with the following data:

Field Name	Value
LrtCommon.ResourceName	Corpus PAAU 92
LrtCommon.Institute	Institut Universita
LrtCommon.ResourceType	Written Corpus
LrtCommon.LanguagesOther	
LrtCommon.Description	The electronic v
LrtCommon.ContactPerson	Sergi Torner <se
LrtCommon.Format	
LrtCommon.MetadataLink	http://gilmere.up
LrtCommon.Publications	
LrtCommon.ReadilyAvailable	true
LrtCommon.ReferenceLink	http://www.iula.t
LrtDistributionClassification.Non...	false
LrtDistributionClassification.Usa...	false
LrtDistributionClassification.Mod...	false
LrtDistributionClassification.Othe...	
LrtIPR.EthicalReference	
LrtIPR.LegalReference	
LrtIPR.LicenseType	
LrtIPR.Description	

The right window, titled 'LrtCommon.Description in Corpus PAAU 92', shows a detailed view of the 'LrtCommon.Description' field. It contains the following text:

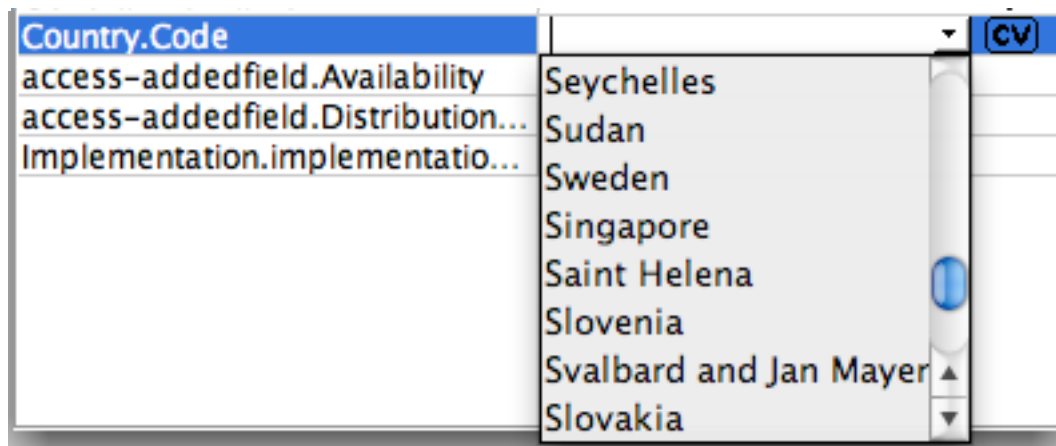
short clarifying description of the resource

The electronic version of the book "Corpus PAAU 1992: Descriptive Studies, Texts and Vocabulary" includes the texts that have been object of analysis in this project as well as the vocabulary lists that make up the Corpus 92.

At the bottom of the right window, there are 'Previous' and 'Next' buttons.

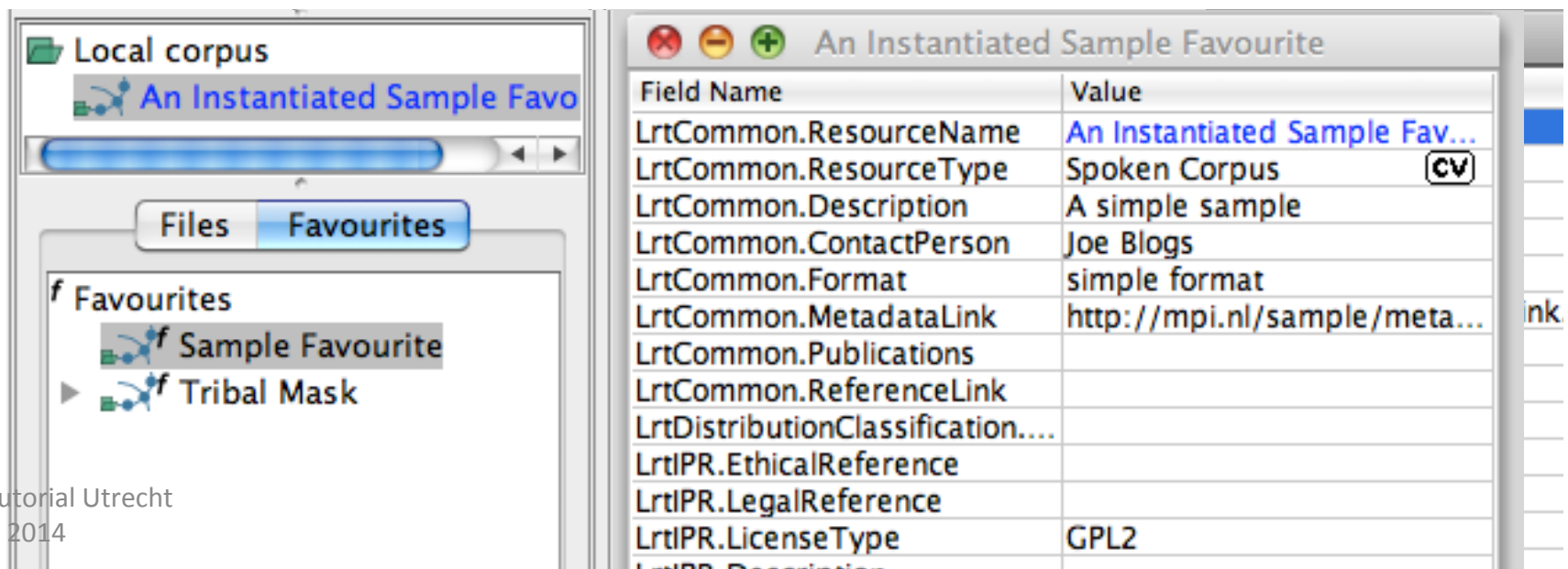
Controlled Vocabularies

- Clarin vocabularies are read from the schema file.
- IMDI vocabularies are read from an XML file.
- Both are provided as dropdown lists in the table when editing.



Using Favourites

- Favourites are snippets of prefilled metadata instances
- New metadata instances can be created from them, leaving only the specific details to be edited
- Either the entire favourite can be used or just the desired sections added to existing metadata

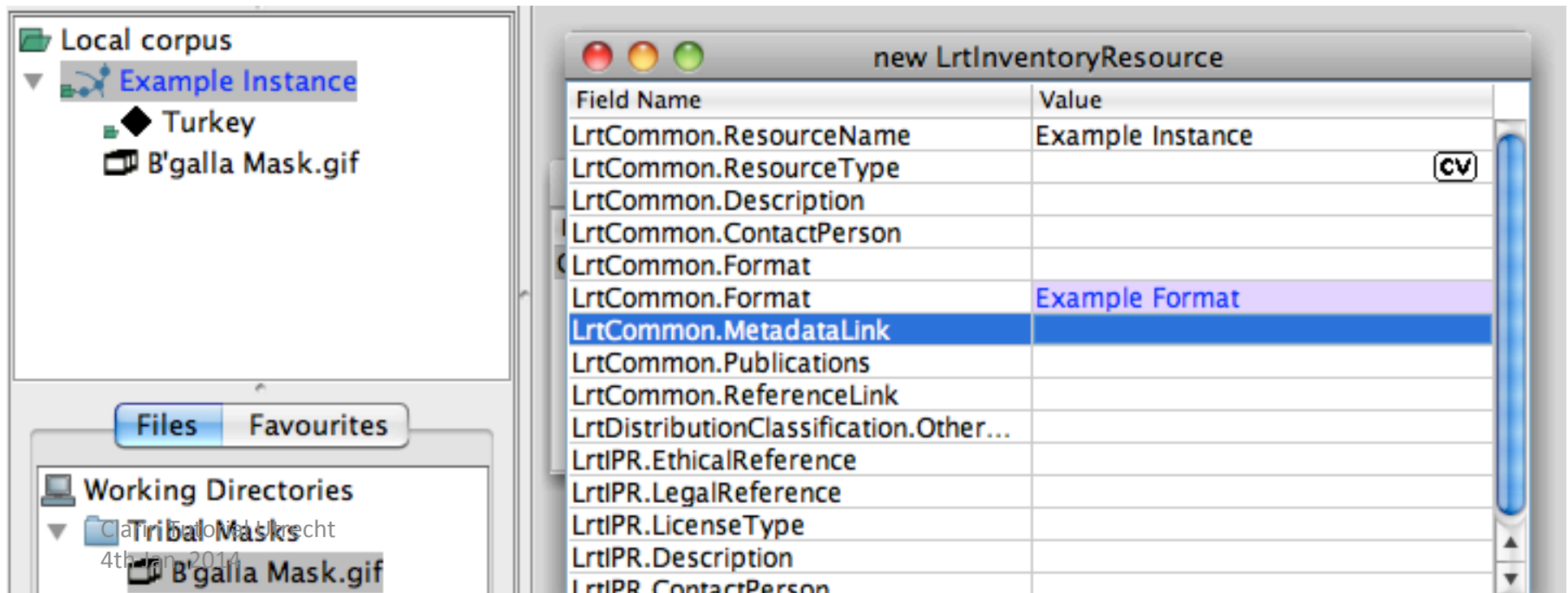


The screenshot shows two windows from the CLARIN-NL interface. The left window, titled 'Local corpus', has a 'Favourites' tab selected, showing a list of favourites: 'Sample Favourite' and 'Tribal Mask'. The right window, titled 'An Instantiated Sample Favourite', displays a metadata table with the following data:

Field Name	Value
LrtCommon.ResourceName	An Instantiated Sample Fav...
LrtCommon.ResourceType	Spoken Corpus CV
LrtCommon.Description	A simple sample
LrtCommon.ContactPerson	Joe Blogs
LrtCommon.Format	simple format
LrtCommon.MetadataLink	http://mpi.nl/sample/meta...
LrtCommon.Publications	
LrtCommon.ReferenceLink	
LrtDistributionClassification...	
LrtIPR.EthicalReference	
LrtIPR.LegalReference	
LrtIPR.LicenseType	GPL2
LrtIPR.Description	

Creating Metadata for a Resource

- Create a metadata instance
- Add a data file
- Add a metadata subsection
- Add a field



The screenshot displays a software interface for creating metadata. On the left, a file browser shows a 'Local corpus' containing an 'Example Instance' with a file named 'B'galla Mask.gif'. Below the file browser are 'Files' and 'Favourites' tabs, and a 'Working Directories' section showing a folder named 'Tribal Masks' with a file named 'B'galla Mask.gif'.

On the right, a window titled 'new LrtInventoryResource' contains a table with the following fields and values:

Field Name	Value
LrtCommon.ResourceName	Example Instance
LrtCommon.ResourceType	
LrtCommon.Description	
LrtCommon.ContactPerson	
LrtCommon.Format	
LrtCommon.Format	Example Format
LrtCommon.MetadataLink	
LrtCommon.Publications	
LrtCommon.ReferenceLink	
LrtDistributionClassification.Other...	
LrtIPR.EthicalReference	
LrtIPR.LegalReference	
LrtIPR.LicenseType	
LrtIPR.Description	
LrtIPR.ContactPerson	



Searching the Metadata

- Searches in Arbil are tree based.
- Multiple search parameters can be entered.
- The search results are shown in a standard table with the usual editing facilities.

The screenshot shows a search window titled "Search" with a "Sample Corpus" icon. The search criteria are set to "All" for the first column and "Tribal" for the second column. The second search row is set to "or" for the first column and "Instance" for the second column. The search status is "searched: 10/10 found: 4".

	LrtCommon.Re...	toolType	LrtCommon.Resou...	applic...	website	LrtCommon.D...	Gener...	LrtCo.
Example I...	Example Insta...		<multiple values>					
Sample In...	Sample Instance		Spoken Corpus			A simple sam...		Joe Bl.
Tribal Mask		Tribal Mask						
Tribal Ma...		Tribal Mas...						

Clarín Tutorial Utrecht
4th Jan, 2014

Find / Replace and Highlighting Cells

- The table can be searched for specific text.
- Selected cells can have the found text substituted.
- Matching table cells can also be highlighted.

Selection								
	LrtCommon.Res...	toolType	LrtCommon.Resou...	applic...	website	LrtCommon.D...	Gener...	LrtCo
	Example M... Example Metad...		<multiple values>					
	Sample Me... Sample Metadata		Spoken Corpus			A simple sam...		Joe Bl
	Tribal Mask	Tribal Mask						
	Tribal Mask	Tribal Mask						
	Tribal Mask	Tribal Mask						
	Tribal Mask	Tribal Mask						
	Tribal Mas...	Tribal Mask 2						

Clarín Tutorial Utrecht
4th Jan, 2014



Installing Arbil

- There is a link to ARBIL on the MPI website
<http://tla.mpi.nl/tools/tla-tools/arbil/>
- Providing you already have Java installed the webstart version is the fastest way to start
- Alternately there are installers for Windows, Mac and Ubuntu (Debian).
- The manual and user guide are also available for download on the same page

The screenshot shows the website for 'The Language Archive' at the Max Planck Institute for Psycholinguistics. The page is titled 'Arbil' and describes it as a general metadata editor, browser & organizer tool for IMDI, CMDI and similar metadata formats. It provides links for 'More information...', 'Release history...', 'System requirements...', and 'How to cite Arbil...'. Under the 'Download' section, it offers the 'Latest stable version' with options to 'Run via webstart', 'Download the Windows installer', 'Download the Debian Linux package', and 'Download the Mac installer'. There is also a section for 'Other versions' with a 'Show all versions' link. The right sidebar contains a search bar, user login information (logged in as petwit@mpi.nl), and contact details for The Language Archive, including street and mailing addresses in Nijmegen, The Netherlands, and social media links for Facebook, Twitter, and RSS.